

Building a News Recommender

Sarabjot Singh Anand, ssanand@tatradata.com

Problem Statement

You have been recruited as data scientists by a start-up, JhakaasNewsVala, based out of Mumbai. The company is developing an app that promises to deliver a unique news experience to its app users.

The company has identified its target market as working professionals in the age group 21-40.

Recognising the fact that retention (defined here as a visit after the first visit) is a huge issue for apps, they understand the need to make an impact on the first visit itself. The problem however is that they know nothing about the user interests or demographics at the time to personalise the news feed to them.

The company has acquired a corpus of news stories.

The real estate available for providing news stories (the mobile phone's screen) is limited and so without a scroll, only 10 stories can be displayed. Statistics show that the number of users scrolling beyond the first set of stories drops off very quickly unless a story on the first page catches the user's eye (that is, results in a clickthrough).

You have been tasked with the job of building two intelligent bots.

1. The article recommender: This bot selects articles to serve a user. Inputs to the bot are the corpus of new articles and a user profile if available.
2. The user profiler: Once the user starts consuming news stories, (s)he leaves behind a clickstream of the form below:

UserId	SessionID	ArticleID Served	Article Rank	Click	Time Spent (seconds)
1	1	28	1	No	
1	1	66	2	No	
1	1	45	3	Yes	69
1
1	1	16	10	No	
1	2	36	1	Yes	46
1

The bot must extract user interests from such data that can then be used for further personalisation for (her)his news feed.

The ultimate objective is to increase clickthrough and the frequency with which the user opens the app to consume stories. However, the objective in the first visit is to:

- Reduce bias in data collection (Example Bias: Stories that get served often and ranked higher, have a higher likelihood of being consumed (obtaining a clickthrough))
- Learn as much as possible about the users on their first visit
- Maximise coverage of the news corpus

Having learned about the trade-off between Exploiting what you know and Exploring the space for what you don't, how would you implement a strategy for the same within the project?

Team Size

4 students per team